

**SYSTEMS AND METHODS FOR UPDATING A FRAME
BUFFER BASED ON ARBITRARY GRAPHICS CALLS**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is related by subject matter to the inventions disclosed in the following commonly assigned applications: U.S. Patent Application No. (not yet assigned) (Atty. Docket No. MSFT-1787), filed on even date herewith, entitled "SYSTEMS AND METHODS FOR EFFICIENTLY DISPLAYING GRAPHICS ON A DISPLAY DEVICE REGARDLESS OF PHYSICAL ORIENTATION"; and U.S. Patent Application No. (not yet assigned) (Atty. Docket No. MSFT-1794), filed on even date herewith, entitled "SYSTEMS AND METHODS FOR EFFICIENTLY UPDATING COMPLEX GRAPHICS IN A COMPUTER SYSTEM BY BY-PASSING THE GRAPHICAL PROCESSING UNIT AND RENDERING GRAPHICS IN MAIN MEMORY".

FIELD OF THE INVENTION

[0002] The present invention relates generally to the field of computer graphics, and more particularly to the efficient generation and updating of computer graphics in a computer frame buffer for display on a display device.

BACKGROUND OF THE INVENTION

[0003] There are many approaches to updating graphics on a display device. One classic method, although rarely used, is the brute force approach where changes to the display graphic are rendered by the processor to memory, and the entire updated graphic is then copied directly to the frame buffer for display. However, this method is extremely inefficient because every pixel of the display device is updated in the frame buffer whether the data for that pixel has changed or not, and the processing resources consumed by this approach are enormous.

[0004] A second method for updating graphics on a display device is for the processor to use a revision list to track in memory each pixel that is changed, and then copy only the updated pixels from memory to the frame buffer. This approach has the advantage of copying to

the frame buffer data pertaining only to those pixels which have changed; however, this approach is also resource intensive in regard to the memory necessary for maintaining the revision list which, in the worst case scenario, may require a change to every pixel. This, along with other shortcomings, significantly slows video processing.

[0005] A third method for updating graphics on a display device involves a complex algorithmic approach that analyzes individual revisions and groups them geometrically into small but efficient "revision regions" comprising both "dirty" (changed) pixels as well as "clean" (unchanged) pixels. The regions are then merged together for an update to the frame buffer. However, for complex revisions, such as a curves and other shapes that can only be broken down into a very large number of small rectangular regions, conducting the merge (among other tasks) is very expensive computationally.

[0006] What is needed in the art is a resource-efficient approach to updating graphics on a display device. The present invention addresses these shortcomings.

SUMMARY OF THE INVENTION

[0007] The method for one embodiment of the present invention is to establish the zone grid at system initialization and, thereafter, track which zones have any pixels revised so that, when the time comes to update the display, only the zones requiring revision (that is, those zones in which any pixel has been revised) are copied from shadow memory to the frame buffer for display on the display device. The memory for tracking these zones can be allocated at initialization and held since it is relatively small. As a result, a significant performance gain may be achieved by avoiding the shortcomings of the existing methods in the art notwithstanding the fact that some "clean" pixels in each zone having even a single changed pixel are also rewritten to the frame buffer.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The foregoing summary, as well as the following detailed description of preferred embodiments, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there is shown in the drawings exemplary constructions of the invention; however, the invention is not limited to the specific methods and instrumentalities disclosed. In the drawings:

[0009] Fig. 1 is a block diagram representing a computer system in which aspects of the present invention may be incorporated;

[0010] Fig. 2A is a block diagram illustrating a computer subsystem where graphics are rendered by the central processing unit (CPU) in main memory (RAM);

[0011] Fig. 2B is a block diagram illustrating a computer subsystem where graphics are rendered by a specialized graphical processing unit (GPU) in video memory (VRAM);

[0012] Fig. 2C is a block diagram illustrating the computer subsystems shown in both Fig. 2A and Fig. 2B coexisting on the same computer system;

[0013] Fig. 3 is a block diagram illustrates the display area of a display device divided into a plurality of zones; and

[0014] Fig. 4 is a flow chart illustrating the method for tracking revised zone and updating the display based on these revised zones.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0015] The subject matter is described with specificity to meet statutory requirements. However, the description itself is not intended to limit the scope of this patent. Rather, the inventors have contemplated that the claimed subject matter might also be embodied in other ways, to include different steps or combinations of steps similar to the ones described in this document, in conjunction with other present or future technologies. Moreover, although the term “step” may be used herein to connote different elements of methods employed, the term should not be interpreted as implying any particular order among or between various steps herein disclosed unless and except when the order of individual steps is explicitly described.

Computer Environment

[0016] Numerous embodiments of the present invention may execute on a computer. Fig. 1 and the following discussion is intended to provide a brief general description of a suitable computing environment in which the invention may be implemented. Although not required, the invention will be described in the general context of computer executable instructions, such as program modules, being executed by a computer, such as a client workstation or a server. Generally, program modules include routines, programs, objects, components, data structures and the like that perform particular tasks or implement particular abstract data types. Moreover,

those skilled in the art will appreciate that the invention may be practiced with other computer system configurations, including hand held devices, multi processor systems, microprocessor based or programmable consumer electronics, network PCs, minicomputers, mainframe computers and the like. The invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[0017] As shown in Fig. 1, an exemplary general purpose computing system includes a conventional personal computer 20 or the like, including a processing unit 21, a system memory 22, and a system bus 23 that couples various system components including the system memory to the processing unit 21. The system bus 23 may be any of several types of bus structures including a memory bus or memory controller, a peripheral bus, and a local bus using any of a variety of bus architectures. The system memory includes read only memory (ROM) 24 and random access memory (RAM) 25. A basic input/output system 26 (BIOS), containing the basic routines that help to transfer information between elements within the personal computer 20, such as during start up, is stored in ROM 24. The personal computer 20 may further include a hard disk drive 27 for reading from and writing to a hard disk, not shown, a magnetic disk drive 28 for reading from or writing to a removable magnetic disk 29, and an optical disk drive 30 for reading from or writing to a removable optical disk 31 such as a CD ROM or other optical media. The hard disk drive 27, magnetic disk drive 28, and optical disk drive 30 are connected to the system bus 23 by a hard disk drive interface 32, a magnetic disk drive interface 33, and an optical drive interface 34, respectively. The drives and their associated computer readable media provide non volatile storage of computer readable instructions, data structures, program modules and other data for the personal computer 20. Although the exemplary environment described herein employs a hard disk, a removable magnetic disk 29 and a removable optical disk 31, it should be appreciated by those skilled in the art that other types of computer readable media which can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, random access memories (RAMs), read only memories (ROMs) and the like may also be used in the exemplary operating environment.

[0018] A number of program modules may be stored on the hard disk, magnetic disk 29, optical disk 31, ROM 24 or RAM 25, including an operating system 35, one or more application programs 36, other program modules 37 and program data 38. A user may enter commands and information into the personal computer 20 through input devices such as a keyboard 40 and pointing device 42. Other input devices (not shown) may include a microphone, joystick, game pad, satellite disk, scanner or the like. These and other input devices are often connected to the processing unit 21 through a serial port interface 46 that is coupled to the system bus, but may be connected by other interfaces, such as a parallel port, game port or universal serial bus (USB). A monitor 47 or other type of display device is also connected to the system bus 23 via an interface, such as a video adapter 48. In addition to the monitor 47, personal computers typically include other peripheral output devices (not shown), such as speakers and printers. The exemplary system of Fig. 1 also includes a host adapter 55, Small Computer System Interface (SCSI) bus 56, and an external storage device 62 connected to the SCSI bus 56.

[0019] The personal computer 20 may operate in a networked environment using logical connections to one or more remote computers, such as a remote computer 49. The remote computer 49 may be another personal computer, a server, a router, a network PC, a peer device or other common network node, and typically includes many or all of the elements described above relative to the personal computer 20, although only a memory storage device 50 has been illustrated in Fig. 1. The logical connections depicted in Fig. 1 include a local area network (LAN) 51 and a wide area network (WAN) 52. Such networking environments are commonplace in offices, enterprise wide computer networks, intranets and the Internet.

[0020] When used in a LAN networking environment, the personal computer 20 is connected to the LAN 51 through a network interface or adapter 53. When used in a WAN networking environment, the personal computer 20 typically includes a modem 54 or other means for establishing communications over the wide area network 52, such as the Internet. The modem 54, which may be internal or external, is connected to the system bus 23 via the serial port interface 46. In a networked environment, program modules depicted relative to the personal computer 20, or portions thereof, may be stored in the remote memory storage device. It will be appreciated that the network connections shown are exemplary and other means of establishing a communications link between the computers may be used.

[0021] While it is envisioned that numerous embodiments of the present invention are particularly well-suited for computerized systems, nothing in this document is intended to limit the invention to such embodiments. On the contrary, as used herein the term “computer system” is intended to encompass any and all devices capable of storing and processing information and/or capable of using the stored information to control the behavior or execution of the device itself, regardless of whether such devices are electronic, mechanical, logical, or virtual in nature.

Graphics Processing

[0022] Figs. 2A, 2B, and 2C are block diagrams illustrating the various elements of two general computer system that comprise typical graphics processing subsystems with which various embodiments of the present invention may be utilized. Fig. 2A illustrates a computer subsystem where graphics are rendered by the central processing unit (CPU) in main memory. Fig. 2B illustrates a computer subsystem where graphics are rendered by a specialized graphical processing unit (GPU) in video memory. Fig. 2C illustrates the computer subsystems shown in both Fig. 2A and Fig. 2B coexisting on the same computer system.

[0023] In each system, a graphics processing subsystem comprises a central processing unit 21' that, in turn, comprises a core processor 214 having an on-chip L1 cache (not shown) and is further directly connected to an L2 cache 212. As well-known and appreciated by those of skill in the art, the CPU 21' accessing data and instructions in cache memory is much more efficient than having to access data and instructions in random access memory (RAM 25, referring to Fig. 1). The L1 cache is usually built onto the microprocessor chip itself, e.g., the Intel MMX microprocessor comes with a 32KB L1 cache. The L2 cache 212, on the other hand, is usually on a separate chip (or possibly on an expansion card) but can still be accessed more quickly than RAM, and is usually larger than the L1 cache, e.g., one megabyte is a common size for a L2 cache.

[0024] For each system in these examples, and in contrast to the typical system illustrated in Fig. 1, the CPU 21' herein is then connected to an accelerated graphics port (AGP) 230. The AGP provides a point-to-point connection between the CPU 21', the system random access memory (RAM) 25', and graphics card 240, and further connects these three components to other input/output (I/O) devices 232—such as a hard disk drive 32, magnetic disk drive 34, network 53, and/or peripheral devices illustrated in Fig. 1—via a traditional system bus

such as a PCI bus 23'. The presence of AGP also denotes that the computer system favors a system-to-video flow of data traffic—that is, that more traffic will flow from the CPU 21' and its system RAM 25' to the graphics card 240 than vice versa—because the AGP is typically designed to allow up to four times as much data to flow to the graphics card 240 than back from the graphics card 240.

[0025] Also common to Figs. 2A, 2B, and 2C is the frame buffer 246 (on the graphics card 240) which is directly connected to the display device 47'. As well-known and appreciated by those of skill in the art, the frame buffer is typically dual-ported memory that allows a processor (the GPU 242 in Fig. 2B or the CPU 21' in Fig. 2A, as the case may be) to write a new (or revised) image to the frame buffer while the display device 47' is simultaneously reading from the frame buffer to refresh the current display content.

[0026] In the subsystem of Fig. 2A (and as further reflected in Fig. 2C), the system RAM 25' may comprise the operating system 35', a video driver 224, and video shadow memory (VSM) 222. The VSM, which is a mirror image of the frame buffer 246 on the graphics card 240, is the location in RAM 25' where the CPU 21' constructs graphic images and revisions to current graphics, and from where the CPU 21' copies graphic images to the frame buffer 246 of the graphics card 240 via the AGP 230. Using this subsystem, certain embodiments of the present invention may be directly executed by the CPU 21' and the RAM 25'.

[0027] In the subsystem of Fig. 2B (and as further reflected in Fig. 2C), the graphics card 240 may comprise a graphics processing unit (GPU) 242, video random access memory (VRAM) 244, and the frame buffer 246. The VRAM 244 further comprises a VRAM shadow memory (VRAMSM) 248. The GPU 242 and VRAMSM 248 essentially mirror the functionality of the CPU 21' and the VSM 222 of Fig. 2A for the specific purposes of rendering video. By offloading this functionality to the graphics card 240, the CPU 21' and VSM 222 are freed from these tasks. For this reason, certain embodiments of the present invention may be directly executed by the components of the graphics card 240 as herein described.

[0028] Again, Fig. 2C shows both of the subsystems of Figs. 2A and 2B co-existing within a single computer system where the computer system itself ostensibly has the ability to utilize either subsystem to execute corresponding embodiments of the present invention.

[0029] As previously discussed earlier herein, there are many approaches to updating graphics on a display device. With the brute force approach, and in reference to Fig. 2C,

changes to the display graphic are rendered by processor (by the CPU 21' or the GPU 242) to memory (VSM 222 or VRAMSM 248). The entire updated graphic is then copied from memory directly to the frame buffer for display. However, this method is extremely inefficient because every pixel of the display device is updated in the frame buffer whether the data for that pixel has changed or not. Moreover, at four bytes per pixel (for 32-bit true color) on a 1024x768 display device, each such update requires copying more than 3MB of graphics data to the frame buffer, and thus the processing resources consumed by this approach are enormous.

[0030] A second known method for updating graphics on a display device is for the processor (the CPU 21' or the GPU 242) to use a revision list to track in memory (VSM 222 or VRAMSM 248) each pixel that is changed, and then copy only the updated pixels from memory to the frame buffer. This approach has the advantage of copying to the frame buffer data pertaining only to those pixels which have changed; however, this approach is also resource intensive in regard to the memory necessary for maintaining the revision list which, in the worst case scenario, may require a change to every pixel. At four bytes per pixel (for 32-bit true color) on a 1024x768 display device (having 1024 pixels per row and 768 pixels per column on the display device), this method requires nearly three megabytes of memory for the revision list. Since this amount of memory typically cannot be allocated and held by the system because of the negative impact such exclusive use of this memory would have on the processing speed of other, unrelated applications, this memory must be allocated (and, thereafter, released) real-time as the revisions are made. However, this amount of memory may not always be available for immediate use. Consequently, the graphic rendering software must have error-handling routines for out-of-memory conditions that might arise when required memory cannot be allocated. Altogether these shortcomings significantly slow video processing using this method.

[0031] A third method for updating graphics on a display device involves a complex algorithmic approach that analyzes individual revisions and groups them geometrically into small but efficient "revision regions" comprising both "dirty" pixels (pixels that have been changed) as well as "clean" pixels (that are unchanged). For efficiency, these regions are dynamically created and tracked in memory by various methods (e.g., by tracking starting point, number of horizontal pixels, and number of vertical pixels to rewrite) and are then merged together for an update to the frame buffer. However, for complex revisions, such as a curves and other shapes that can only be broken down into a very large number of small rectangular regions,

the computational cost of determining the region size, shape, and location; dynamically allocating memory to track same (and releasing this memory when complete); and conducting the merge of revised regions is altogether very expensive computationally.

[0032] To address these shortcomings, in one embodiment of the present invention, the display area of the display device 47' (and the corresponding frame buffer 246n and/or shadow memories, VSM 222 and/or VRAMSM 248 respectively in Fig. 2C) is divided into a plurality of "zones" as illustrated in Fig. 3. As shown in this figure, the graphical display area of a 1024x768 pixel display device 47' may be divided into 1024 zones forming a 32x32 "zone grid" 302. Each zone (e.g., zone 304) comprises 32x24 pixels (e.g., pixel 306)—that is, each zone has the same dimensions and number of pixels as the other zones. For this embodiment, the width and the height of each zone are each exactly $1/32^{\text{nd}}$ of the width and of the height of the display area of the display unit, and thus the number of zones vertically aligned on the display device is equal to the number of zones horizontally aligned on the display device (thereby forming a "square" zone grid). Moreover, in this particular embodiment, the zones are predefined at startup and are static (do not change).

[0033] In alternative embodiments of the present invention, the zones may be established at some time other than startup (not predetermined), and the zones may be dynamic based on algorithms employed to determine the most optimal zone size for any particular use (e.g., larger zones for text-based applications, smaller zones for applications that render detailed graphics objects) when the increased overhead necessary may be justified. Likewise, other alternative embodiments may not comprise a square zone grid but, instead, comprise a rectangular grid when the number of vertical zones is greater or less than the number of horizontal zones.

[0034] The method for one embodiment of the present invention, as illustrated in Fig. 4 is to establish the zone grid at system initialization 402 and, thereafter, track which zones (e.g., zone 404) have any pixel revised so that, when the time comes to update the display, only the zones requiring revision are copied from shadow memory to the frame buffer 406. The method of this embodiment significantly reduces the overhead required to track the changes in the zone as only the starting point of each zone need be listed as the number of horizontal and vertical pixels for each zone is fixed, and the memory for tracking these zones can be allocated at initialization and held since it is relatively small. As a result, a significant performance gain may

be achieved by avoiding the shortcomings of the existing methods in the art notwithstanding the fact that some "clean" pixels in each zone having even a single changed pixel are also rewritten to the frame buffer.

[0035] The foregoing method is particularly effective computer systems utilizing text-enhancement technologies (TETs) such as Microsoft's ClearType™. ClearType™ is a "sub-pixel anti-aliased," a special type of TET software that dramatically improves the readability of text on LCDs (Liquid Crystal Displays), including without limitation laptop screens, Pocket PC screens, and flat panel monitors. ClearType™ enables the words on a display monitor to appear almost as sharp and clear as those printed on a piece of paper. This particular TET works by accessing the individual vertical color stripe elements (sub-pixels) in every pixel of an LCD screen. Prior to ClearType™, the smallest level of detail that a computer could display was a single pixel, but this TET displays features of text as small as a fraction of a pixel in width. This extra resolution increases the sharpness of the tiny details in text display, making it much easier to read over long durations. However, in operation this TET necessarily renders a very large number of graphic revisions to more clearly display the text, and these revisions are most effectively and efficiently rendered using the method described for the present embodiment.

Conclusion

[0036] The various system, methods, and techniques described herein may be implemented with hardware or software or, where appropriate, with a combination of both. Thus, the methods and apparatus of the present invention, or certain aspects or portions thereof, may take the form of program code (i.e., instructions) embodied in tangible media, such as floppy diskettes, CD-ROMs, hard drives, or any other machine-readable storage medium, wherein, when the program code is loaded into and executed by a machine, such as a computer, the machine becomes an apparatus for practicing the invention. In the case of program code execution on programmable computers, the computer will generally include a processor, a storage medium readable by the processor (including volatile and non-volatile memory and/or storage elements), at least one input device, and at least one output device. One or more programs are preferably implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the program(s) can be

implemented in assembly or machine language, if desired. In any case, the language may be a compiled or interpreted language, and combined with hardware implementations.

[0037] The methods and apparatus of the present invention may also be embodied in the form of program code that is transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via any other form of transmission, wherein, when the program code is received and loaded into and executed by a machine, such as an EPROM, a gate array, a programmable logic device (PLD), a client computer, a video recorder or the like, the machine becomes an apparatus for practicing the invention. When implemented on a general-purpose processor, the program code combines with the processor to provide a unique apparatus that operates to perform the indexing functionality of the present invention.

[0038] While the present invention has been described in connection with the preferred embodiments of the various figures, it is to be understood that other similar embodiments may be used or modifications and additions may be made to the described embodiment for performing the same function of the present invention without deviating there from. For example, while exemplary embodiments of the invention are described in the context of digital devices emulating the functionality of personal computers, one skilled in the art will recognize that the present invention is not limited to such digital devices, as described in the present application may apply to any number of existing or emerging computing devices or environments, such as a gaming console, handheld computer, portable computer, etc. whether wired or wireless, and may be applied to any number of such computing devices connected via a communications network, and interacting across the network. Furthermore, it should be emphasized that a variety of computer platforms, including handheld device operating systems and other application specific hardware/software interface systems, are herein contemplated, especially as the number of wireless networked devices continues to proliferate. Therefore, the present invention should not be limited to any single embodiment, but rather construed in breadth and scope in accordance with the appended claims.

[Remainder of Page Intentionally Left Blank]